Publishing Community Efforts and Solutions to Mitigate the Risks Sci-Hub Poses to Researchers, Librarians, and Publishers

Sari Frances
IEEE, s.frances@ieee.org

Juan P. Denzer
SUNY Oswego, juan.denzer@oswego.edu

Don Hamparian
OCLC, hamparid@oclc.org

Sari Frances, Juan P. Denzer, and Don Hamparian, "Publishing Community Efforts and Solutions to Mitigate the Risks Sci-Hub Poses to Researchers, Librarians, and Publishers" (2018). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/https://doi.org/10.5703/1288284317045

Publishing Community Efforts and Solutions to Mitigate the Risks Sci-Hub Poses to Researchers, Librarians, and Publishers

Sari Frances, Manager of Digital License Compliance, IEEE, s.frances@ieee.org

Juan P. Denzer, Discovery Services Librarian, SUNY Oswego, Juan.denzer@oswego.edu

Don Hamparian, Product Manager, OCLC, hamparid@oclc.org

Abstract

Sci-Hub has been referred to as the "Robin Hood" of science, but in reality, it is not. Sci-Hub is a disruption to the entire scholarly publishing research cycle. Over the last three years, the amount of licensed e-content that has been illegally obtained by Sci-Hub has grown significantly. This content has been acquired through stolen institutional staff and student credentials. Acquiring and misappropriating these credentials creates serious risks for an institution's systems and users as well as publishers. What can libraries and publishers do to minimize or eliminate these infractions? This discussion about the collective efforts of publishers, libraries, and other organizations will cover how to protect electronic resources, personal data, and adopting best practices in order to better defend from cyberattacks that compromise our organizations. We will discuss how these attacks can occur and steps you can take to protect your library. While, unfortunately, there is no one single solution for this problem, we will also look at a case study of a library that successfully implemented some of these practices to combat the cyberattacks. Through this we will demonstrate that together we can reduce the risks facing researchers, librarians, and publishers created by Sci-Hub.

How Does Sci-Hub Acquire Content? Why Should You Care?

Ultimately, Sci-Hub's illegal activities harm learned societies that are reliant on subscription income to support their important work; it is a threat to the scholarly communications ecosystem, the sustainability of high-quality journals, as well as the ability to invest in new journals and fields. Sci-Hub has no incentive to ensure the accuracy of the research articles being accessed, and its continued operation poses a threat to the accuracy of the scientific record (Page, 2018). This statement given by Elsevier and Springer Nature is one of the main reasons why publishers, including IEEE and other not-for-profits, are taking action and not taking a back seat.

Sci-Hub steals scholarly content predominantly by using stolen user credentials acquired through phishing attacks or other nefarious means. This enables them to illegally access university networks, steal licensed content, and ultimately make it freely available to the public. But what else do these stolen user credentials allow Sci-Hub to access? What else can they do besides steal content? Although we do not have any hard evidence of other types of illegal use (e.g., credit card theft, identity theft), we want to stress the vulnerabilities and risks to personal data that this situation implies.

Not only are publishers and institutions responsible for protecting content and personal data, but the end users who access electronic resources are accountable as well. It is imperative that end users adopt stronger passwords, comply with both university policies and copyright law, and not share credentials or access documents on Sci-Hub. IEEE has found in 99% of these cases where universities have informed IEEE of the interactions with the compromised end user, that the end users don't even realize that their account has been seized. These are not users who are necessarily visiting Sci-Hub or deliberately sharing their network credentials, but rather innocent victims of piracy. Institutions can train staff and students to be aware of these vulnerabilities and the consequences of using Sci-Hub, strengthen security by implementing tools such as dual-factor authentication, and respond to content providers when alerted of such activity. Other recommended steps to secure access can be also be found in an article on the Scholarly Kitchen (https://scholarlykitchen.sspnet.org/2018/04/24/guest-post-technology-law-education-three-pronged-approach-fight-digital-piracy/).

Publishers have and continue to partner with institutions to educate and inform students and faculty about Sci-Hub activity. IEEE has already worked with several universities in the United States, Australia,

https://doi.org/10.5703/1288284317045

and Singapore to develop solutions in order to help them protect the integrity of their networks as well as published content. Many of these compromised institutions have now, with our help, begun to understand that the vulnerabilities lay within their own authentication systems and found ways to mitigate these risks (Frances, 2018).

Publishing Community Efforts

Beginning in 2014, IEEE and other publishers began tracking and alerting institutions about Sci-Hub activity. In early 2017, IEEE began requesting customers to provide data regarding the illicit activity in order to learn more about these attacks on both the university's firewall and IEEE's. This was a mutually beneficial effort to advocate for security for both parties. In discussions with universities, we learned from librarians that temporary suspension of access to the IEEE Xplore platform (a common means to deal with a breach of access) was not an ideal approach. As a result of these discussions, we worked with Juan Denzer from SUNY Binghamton (now at SUNY Oswego) who developed a script to automatically block Sci-Hub activity without having an adverse effect on legitimate users. IEEE staff then modified the script to accommodate institutions using other platforms. Continued feedback has been extremely important in our efforts to provide technical solutions to mitigate Sci-Hub activity. IEEE has a dedicated team to address these concerns and is working on developing additional solutions that are both effective and customer friendly.

Other publisher efforts include the Sci-Hub Executive Steering Group formed in August 2017, consisting of members from 15 different publishers that have come together to address this problem and strategically collaborate on solutions. This includes legal, technical, and educational methods that we hope will disrupt and/or eventually shut down Sci-Hub.

Case Study—Binghamton University

In the fall of 2016, Binghamton University Libraries was contacted by the IEEE manager of digital license compliance. IEEE began to suspend access to their content due to excessive illegal downloads. The block affected patrons, faculty, and staff. The library lack of access meant a loss in spending for a product not being used.

Initially, the library manually located the compromised accounts and took action. The library system administrator would spend hours searching log files to find the accounts that were used in the breach. Each user account was flagged and blocked in the EZproxy server. Patrons were then notified to change their password. Once the EZproxy server was secured, the IEEE manager of digital license compliance was contacted. The library system administrator assured them the server was secure and made a request to reinstate access to IEEE content.

The process became a routine of seek, secure, notify, and reinstate. This periodic plan would happen almost daily. The library was being compromised on a regular basis. Most compromises were between January to February 2017.

By March 2017, the system administrator at Binghamton University determined that the process was using too much of the library's resources. These included hours searching log files, time spent modifying the server, and contacting IEEE to reinstate access. In order to minimize the resources dedicated to combating the issue, the process was automated.

The automated process was achieved by developing a server script to work in three phases: seek, secure, and notify. Seek involved searching the EZproxy log files for a unique flag. This flag was created and embedded by the IEEE technical support team. Locating the flag in the log files is what allowed the library to identify the hacked account. Previously the system administrator had to manually search for the flag. What normally took a large amount of time now took microseconds to complete. This significantly cut the amount of time spent searching log files. The script could also be scheduled to run more often than a manual search.

Secure, the second phase, was the process of securing the EZproxy server. This involved blocking the originating IP address and temporarily blocking the compromised account. The EZproxy server is suspended, so the script can modify the configuration files. This modification adds a block directive to the hacked IP and patron account. The server's service is then restarted. The process is done within seconds. The process is seamless and suspended access to EZproxy is minimal.

The final phase, notify, involved sending an e-mail to the system administrator and IEEE manager of digital license compliance. The system administrator received a time-stamped e-mail with the list of compromised accounts as well as the IP addresses.

Batch	Email Sent	PDF Activity (Start - End)	Status	Number of PDF Downloads
171	12/10/2016 18:32	NA - NA	Closed	0
353	1/22/2017 16:24	2017-01-21 18:40:21 - 2017-01-21 18:40:21	Closed	1
367	1/25/2017 8:07	2017-01-24 23:24:31 - 2017-01-25 01:53:32	Closed	96
397	1/30/2017 15:25	NA - NA	Closed	0
417	2/3/2017 9:59	NA - NA	Closed	0
463	2/12/2017 16:40	2017-02-11 23:31:40 - 2017-02-12 12:56:04	Closed	110
479	2/13/2017 8:09	2017-02-12 22:07:23 - 2017-02-13 00:34:23	Closed	50
482	2/13/2017 13:47	2017-02-13 12:34:29 - 2017-02-13 13:30:04	Closed	29
491	2/15/2017 8:36	2017-02-14 21:33:19 - 2017-02-15 00:02:42	Closed	109
1420	8/29/2017 14:47	NA - NA	Closed	0
1459	9/4/2017 10:45	NA - NA	Closed	0
1516	9/19/2017 9:47	NA - NA	Closed	0
1603	10/3/2017 16:45	NA - NA	Closed	0
1649	10/16/2017 8:44	NA - NA	Closed	0
1683	10/29/2017 12:55	NA - NA	Closed	0
1713	10/30/2017 11:33	NA - NA	Closed	0
1718	10/30/2017 12:54	NA - NA	Closed	0
1720	10/30/2017 13:38	NA - NA	Closed	0
1740	11/7/2017 10:11	NA - NA	Closed	0
1930	12/17/2017 16:15	NA - NA	Closed	0
1953	12/18/2017 10:40	NA - NA	Closed	0

Figure 1. Total IEEE data breaches at Binghamton University for one year.

This allowed the library to forward the list to their campus IT department. The IT department would notify the users on the list to update their password. The second e-mail list sent to IEEE only included the IP addresses. This allowed them to block the IP address on their servers.

Since IEEE was notified automatically, it was not necessary for the library to contact IEEE directly. The list of IP addresses was already an indication that the server was secured.

The server script was a preemptive defense against attacks to the EZproxy server. Since the script ran faster and about every five minutes, any attempts to download content dropped significantly. This is in contrast to the prior process that couldn't keep up with the breaches, resulting in more content being illegally downloaded. Once the script was implemented in August 2017, the illegal downloads went to zero.

Since the implementation of the script, Binghamton University has gone from being the most compromised library to virtually zero compromises. As a result, IEEE technical support developed a similar script for Linux-based servers.

IEEE Efforts: We Can Help

IEEE's Digital Licensing team has been working with libraries and guiding them through the process of tracking and blocking Sci-Hub activity. In addition they provide resources that help prevent Sci-Hub activity either before or after a breach. IEEE has also worked with organizations such as OCLC (www.oclc.org) and PSI (www.publishersolutionsint.com) to develop and implement several freely-available tools (discussed below). They can work with university IT/ITS departments to implement the following tools:

Before a Breach

Option 1: Special code to protect from future breaches

Option 2: Updated Stanza for EZproxy access (prevent blocking)

Option 3: Windows and Linux script solutions (currently working on the next generation)

Option 4: Proactively identify blacklisted/Sci-Hub IP addresses: IP Registry—www.theipregistry.org

These free tools can help protect against future attacks and breaches. Institutions have to decide

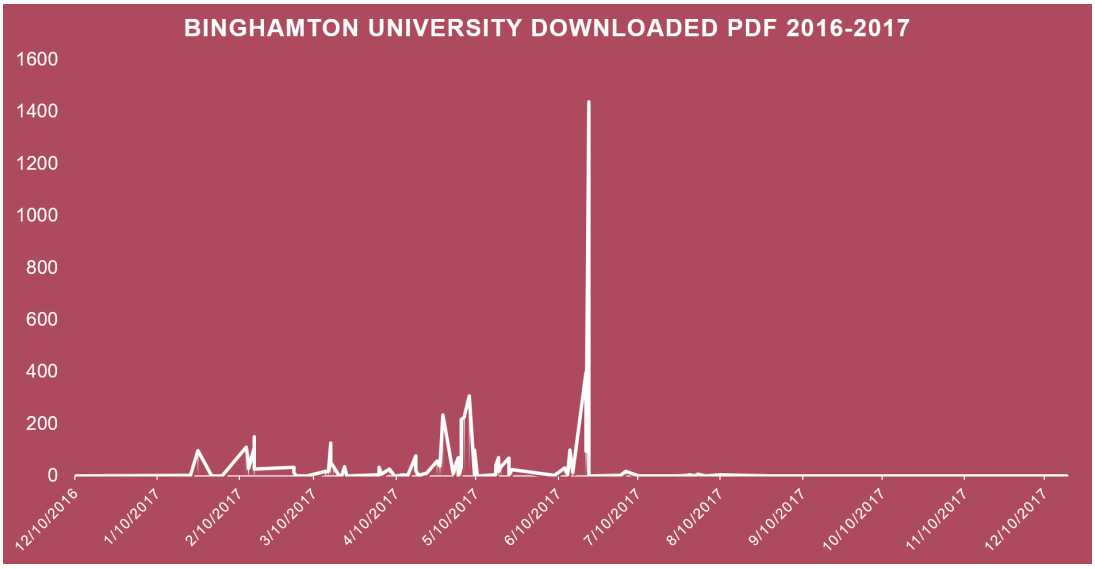

Figure 2. Total daily IEEE data breaches at Binghamton University from December 2016 to December 2017.

what's right for their library and the resources needed to implement some or all of these tools.

After a Breach

- IEEE sends log files that include:
  - Date and time
  - Information on how to remediate
- Collaborate with IT/ITS department
- Alleviate downtime for customers in different time zones (respond within less than 24 hours)

EZproxy and RA21

EZproxy access and authentication software allows your library to deliver e-content simply and effectively. EZproxy facilitates a single sign-in to e-content using existing library-issued credentials, such as a library card number and PIN or username and password. EZproxy interoperates with a number of authentication facilities including Shibboleth (SAML), LDAP, Active Directory, and SIP.

Normally, EZproxy uses credentials (username, password, multifactor if present) supplied by institution identity providers (IdP). RA21 emphasizes using SAML-based IdPs and EZproxy supports this authentication method.

Resource Access for the 21st Century (*RA21*) is a joint *STM* and *NISO* initiative aimed at optimizing protocols across key stakeholder groups, with a goal of facilitating a seamless user experience for consumers of scientific communication. In addition, this comprehensive initiative is working to solve long-standing, complex, and broadly distributed challenges in the areas of network security and user privacy.

It is likely that RA21 will specify updated patterns and/or standards for SAML-based authentication workflows. EZproxy interoperates with RA21-enabled sites today and will be updated to support applicable RA21 patterns and/or standards.

Security Policies to Protect Against Breaches

The majority of successful breaches come from stolen credentials. Credentials can be guessed or phished for (usually via e-mail). It is essential that good credential management practices and policies are defined at your institution (including the library and EZproxy).

The best defense against stolen credentials is multifactor authentication (MFA) where another credential is supplied from a different source such as am SMS message. Many institution identity providers support MFA.

If MFA is not available, then developing and adhering to a set of security management policies oriented around passwords is essential. Since EZproxy is part of the security domain, it should be included in the covered applications of the policy.

The policy should contain:

- Short password lifetimes (less than 90 days)
- Password complexity requirements (not dictionary words, letters, numbers, punctuation required)
- A breach detection process: how to validate that a set of credentials has been used for malicious access
- Frequent exercising of the breach detection process
- Education plan for students, staff, and patrons

The breach detection process should include understanding where breaches can be detected and a written plan for resolution including detection steps, and organizations that need to be involved such as the institution's IT, publishers, and so on.

The education plan's goal is to show students and staff that stealing credentials has a widespread effect in a typical institution environment. Institutions that have a widely implemented single-sign-on (SSO) infrastructure have many systems—not related to library access—that become available when credentials are stolen including financial, scheduling, and course learning systems. Once students and staff realize the personal cost of stolen credentials, they tend to take the security policies more seriously.

EZproxy Configurations to Protect Against Breaches

EZproxy currently has a number of tools to help protect and diagnose malicious access. The EZproxy websiteand Community Center have a number of tools identified. Additionally, these actions are recommended:

- Ensure that EZproxy's log files (audit log, messages.txt, and ezproxy.log) are secured via appropriate operating system permissions.
- Keep and back up at least 180 days of EZproxy log data.
- Use SSL for all authentication interactions and to all content providers that support SSL.
- Keep your server operating system upgraded with the latest vendor patches.
- Keep your server's clock (time) correctly set against Internet standard time servers.
- Upgrade your EZproxy configuration to the latest version. Each version of EZproxy updates a number of security-related features.

Some publishers recommend including the EZproxy Option X-Forwarded-For configuration statement. This statement tells EZproxy to send the IP address of the browser user to the publisher. This information can be helpful to the publisher when malicious usage occurs and helps avoid the publisher shutting down the IP address of the proxy service, denying access to all users.

As described above, it's important to exercise breach detection processes regularly and before a publisher contacts you threatening to turn off access. The detection process for malicious access through EZproxy has two main paths:

- Being able to identify a user session from publisher-supplied information (which normally includes URLs and date/time of access)
- Being able to identify potential compromised users based on user behavior

Both of these processes are documented on the OCLC Support website.

Future versions of EZproxy will continue to add new security-related features including automation of some of the manual steps described here. With a solid password and security policy as well as monitoring EZproxy logs, EZproxy provides a secure access process for accessing STEM content.

Conclusion

The threats of security are a problem for the entire academic and publishing community. As a result of collaborating with publishers, libraries, and other organizations, the efforts to combat Sci-Hub have

successfully minimized the threat, but not completely eliminated it. Institutions that have implemented not only their own scripts and processes to protect themselves from imminent attacks have seen a significant drop in piracy. Therefore, current and future solutions along with education implemented by libraries and publishers will assist in the protection against digital piracy.

References

EZproxy access and authentication software. (2018, December 10). Retrieved from https://www.oclc.org/en/ezproxy.html

Frances, S. (2018, April 24). Technology, law, and education: A three-pronged approach to fight digital privacy. *The Scholarly Kitchen.* Retrieved from scholarlykitchen.sspnet.org: https://scholarlykitchen.sspnet.org/2018/04/24/guest-post-technology-law-education-three-pronged-approach-fight-digital-piracy/

OCLC Community Center. (2018). Retrieved from https://www.oclc.org/community/home.en.html

Page, B. (2018, December 11). Publishers succeed in getting Sci-Hub access blocked in Russia. *thebookseller.* Retrieved from thebookseller.com: https://www.thebookseller.com/news/sci-hub-blocked-russia-following-court-action-publishers-911571

RA21: Resource access for the 21st century. (2018). Retrieved from https://ra21.org/

Secure your EZproxy server. (2018, November 1). Retrieved from https://help.oclc.org/Library_Management/EZproxy/Secure_your_EZproxy_server/020Secure_your_EZproxy_server